PATENT / DOCK. NO. 16356.608 (DC-02942) Customer No. 000027683

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of: Low, et al.

Serial No.: 10/080,815

Filed: February 22, 2002

For: METHOD AND SYSTEMS FOR

SCREENING CHINESE ADDRESS DATA

Group Art Unit: 2644

Examiner: Not yet assigned

RECEIVED

JUN 2 4 2002

Technology Center 2600

CLAIM FOR PRIORITY

Commissioner for Patents Washington, DC 20231

Sir:

Pursuant to 35 U.S.C. § 119 and 37 C.F.R. § 1.55(b), Applicants hereby claim the right of priority based on Chinese Patent Application No. 02101736.0 filed in China on January 17, 2002.

In support of this claim for priority, enclosed is a certified copy of the Chinese priority application.

Respectfully submitted,

Jarnes R. Bell Reg. No. 26,528

6-12-02 HAYNES AND BOONE, LLP 901 Main Street, Suite 3100

Dallas, Texas 75202-3789 Telephone: 512/867-8407 Facsimile: 512/867-8470

A-132873.1

I hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail in an envelope addressed to: Commissioner for Patents, Washington, D.C.

20231





CERTIFIED COPY OF PRIORITY DOCUMENT

明

本证明之附件是向本局提交的下列专利申请副本

日: 2002 01 17

号: 02 1 01736.0

别: 发明

占称: 用于鉴别中文地址数据的方法和系统

人: 戴尔产品有限公司

设计人: 刘罗梦华; 俞欣; 陈志伟; 奚珍明; 李来; 曾顺达

ALEMONIUM SIONANDISIONANDISIONANDISIONANDISIONANDISIONANDISIONANDISIONANDISIONANDISIONANDISIONANDISIONANDISIONA

中华人民共和国 国家知识产权局局长 2 季 川

2002 年 4 月 19 日

1. 一种用于比较两个数据库的计算机应用方法,每个数据库都包含指明地址的中文文本数据项目,该方法包括:

对于每一个数据库,将任何一个并不是预先定义的共同中文语言格式的中文文本数据项目转换成共同格式,能够用多种方式转换成该共同格式的、且至少是在第一数据库中的任何项目都被用所有这些方法来转换成该共同格式:和

比较这些具有共同格式的数据项目,以便识别与该第二数据库中的中文文本数据项目相对应的该第一数据库中的中文文本数据项目。

- 2. 一种如权利要求 1 所述的方法, 其中该共同的中文语言格式是拼音字符(Pin Yin characters)。
- 3. 一种如权利要求1或权利要求2所述的方法,其中该第一数据库是一个订单管理系统数据库,并且包括含有相应的运输和/或订货地址的数据项目。
- 4. 一种如权利要求 1 至 3 中的任一项所述的方法,其中该第一数据库的中文文本数据项目使用简体汉语字符(Simplified Mandarin characters)。
- 5. 一种如前述的权利要求中的任一项所述的方法,其中该第二数据库是一个由第二方所颁发的第三方列表。
- 6. 一种如前述的权利要求中的任一项所述的方法, 其中不包括位于至少一个所指定中国领土之内地址的该第二数据库数据项目将不被转换成该共同格式。
- 7. 一种如前述的权利要求中的任一项所述的方法,其中通过对具有该共同格式中文文本数据项目的比较,若发现它们之间相匹配,则

1

编辑该第一数据库的相应项目。

- 8. 一种用于比较两个数据库的计算机系统,其中每个数据库包括 指明地址的中文文本数据项目,该计算机系统含有:
- 一第一转换单元,用来将第一数据库的中文文本数据项目转换成一预先定义的共同中文语言格式,并且能够用多种方法将它们转换成共同格式的第一数据库中的任意项目被用所有这些方法来转换以便生成具有共同格式的项目;
- 一第二转换单元,用来将该第二数据库的中文文本数据项目转换 成该共同中文语言格式: 和
- 一比较单元,用于比较被转换的数据项目,以便识别与该第二数据库中的中文文本数据项目相对应的该第一数据库中的中文文本数据项目。
- 9. 一种如权利要求 8 所述的系统, 其中该共同中文语言格式是拼音字符。
- 10. 一种如权利要求 8 或权利要求 9 所述的系统,其中该第一数据库是一个订单管理系统数据库,并且包括含有相应的运输和/或订货地址的中文文本数据项目。
- 11. 一种如权利要求 8 至 10 中的任一项所述的系统,其中该第一数据库的中文文本数据项目使用简体汉语字符。
- 12. 一种如权利要求 8 至 11 中的任一项所述的系统,其中该第二数据库是一个由第二方所颁发的第三方列表。
- 13. 一种如权利要求8至11中的任一项所述的系统,其中该第二转换单元并不转换那些不包含位于至少一个所指定的中国领土之内地址的该第二数据库数据项目。
 - 14. 一种如权利要求 8 至 13 中的任一项所述的系统, 其中通过检

~

测该具有共同格式的中文文本数据项目之间的相匹配情况,该比较单元用来操作输出装置以发送信号,因而使得该第一数据库相应项目被编辑。

15. 一种如权利要求 8 至 13 中的任一项所述的系统,其中通过检测具有该共同格式的中文文本数据项目之间的相匹配情况,该比较单元用来编辑该第一数据库中的相应项目。

用于鉴别中文地址数据的方法和系统

技术领域

本发明涉及用来比较两个中文项的数据库的方法和系统,尤其是,本发明用以将诸如个人和/或机构地址的数据项目进行比较。

背景技术

目前已存在多种书写中文文本的标准,除了传统的中文字符集(它仍然在台湾和香港这些地区广泛地使用),中华人民共和国是用简化的汉字来书写文本,另外中文可以被译成像"拼音字符"这样的罗马字母,或者是通过其他的系统,比如被定义成 ALA-LC 的罗马字母表的系统。

各种标准之间的转换是很常见的,例如一个传统的订单管理系统(SMARTS)需要通过使用拼音字符来键入订货和运输地址,然后该拼音字符被转换成存储于 SMARTS 数据库中的双字节简体汉字字符。注意,尽管如此,并非所有的转换都是很明确的,比如一单个的简体汉字字符可以对应(用拼音)几个罗马字母集。类似地,一单个罗马字母(用拼音)集也可能会对应多个简体汉字字符,这些简体汉字字符将会有不同的意义。

既然不同数据库中的中文字符可以用不同标准来被存储,因而比较不同数据库中的相应项目是一个困难的过程,比如一第二方已经颁布了"被拒绝方列表"("DPL"),并且与列表中的所有方进行的交易都将被特殊处理。该列表只是用英文来发布(即,传统的英文文字和译自中文文字的罗马字母的混合体)并且没有任何迹象表明将来它会被翻译成简体汉语字符。因此使用存储于诸如 SMARTS 这样的订单管理系统中的名字来比较列表是很困难的。

比较这两个列表的困难会引起一种可能,它会让产品提供商错误 地将产品提供给列于 DPL 表上的当事方,而未对该交易进行特殊处理。 这样的错误可能对公司的利益造成损害。

发明内容

本发明试图解决上述问题,尤其是提供用来比较两个数据库的方法和系统,该两种数据库都包括中文文本数据项目比如个人或机构这样实体的地址,并且两种数据库对于中文文本数据项目采用不同的中文书写系统。

总的来说,本发明提出两种数据库的中文文本项都被转换成一共同的标准语言,尤其是拼音拼写标准(Pin Yin Transliteration standard)。在转换过程中,任何可以用多种方法转换的项目都被用这些方法中的进行转换,于是这两个被转换数据库中的项目即可以进行比较。

特别地,本发明的第一方面在于用来比较两个数据库的计算机应用方法,该两数据库的每一个都包括标明地址的中文文本数据项目,该方法包括:

对于每一个数据库,将任何一个并不是预先定义的共同中文语言格式的中文文本数据项目转换成共同格式,能够用多种方式转换成该共同格式的、且至少是在第一数据库中的任何项目都被用所有这些方法来转换成该共同格式:和

比较这些具有共同格式的数据项目,以便识别与该第二数据库中的中文文本数据项目相对应的该第一数据库中的中文文本数据项目。

本发明的第二方面在于提出了一种用于比较两个数据库的一种计算机系统,每个数据库都包括指明地址的中文文本数据项目,该计算机系统包括:

- 一第一转换单元,用来将第一数据库的中文文本数据项目转换成一预先定义的共同中文语言格式,并且能够用多种方法将它们转换成共同格式的第一数据库中的任意项目被用所有这些方法来转换以便生成具有共同格式的项目;
- 一第二转换单元,用来将该第二数据库的中文文本数据项目转换 成该共同中文语言格式;和
- 一比较单元,用于比较被转换的数据项目,以便识别与该第二数据库中的中文文本数据项目相对应的该第一数据库中的中文文本数据项目。

注意如果该第二数据库中的数据项目已经是共同格式,那么可以省略该第二转换单元。

该共同中文语言格式优先采用拼音字符 (Pin Yin Characters),该第一数据库可以是一订单管理系统,它具有包括运输和/或提货地址的数据项目,第一数据库中的中文文本数据项目可能是简体汉语字符。该第二数据库可能是英文或者是传统的英文文字和拼音的结合体,比如,该第二数据库可以是一些或者是全部的由由第二方所颁发的第三方列表。

"中文文本数据项目"可以被定义成中国语言的项目,比如说汉字,可选择地或附加地,"中文文本数据项目"可以被定义成包括或者是含有与位于指定的中国境内的地址有关的数据项目,比如中华人民共和国和/或可选的任何其他中文语言通用的领土内以便订货和/或运输(尤其是使用简体汉字字符地区)。

注意,除了中文文本数据项目之外,任何一个数据库可能包括非中文文本数据项目的项目,比如该订单管理系统可以包括与中国没有任何关系的方面的数据。类似地,该第二数据库(比如说是一些或者是全部的 DPL 方的情况)包括验证其实体的地址不位于指定的中国境内的项目。优先地,在任何一种情况下该转换过程仅仅将每一个数据库中的中文项目进行转换,并且只有该比较才确定该第一数据库的被转换项目是否与该第二数据库的被转换项目相对应。

附图说明

与所描述的本发明的实例相关的本发明的另外的优点和特征将要被讨论,仅仅作为一个例子,参考下列附图,它包括:

图 1 是用来显示作为本发明的具体实例的方法的方框图;

图 2 是作为本发明的具体实例的系统的结构方框图,它用来实现图 1 的方法;

图 3 显示了由图 2 中的系统产生的一个窗口,它被用来形成由拼音字符所得的简体汉语字符地址:

图 4 是由图 2 的系统所产生的一个窗口,它显示了存储于该系统

I∕

中的用简体汉语字符表示的地址:

图 5 由图 5(a)至 5(c)所组成,它显示了图 1 中的将简体汉语字符转换成拼音字符的方法的步骤;

图 6 是图 2 的系统所产生的一个窗口,它显示了两个数据库进行比较的结果。

具体实施方式

图 1 显示了依据本发明的具体实施例的方法的步骤,它用来将潜在的接受物件方的地址与至少是发展方列表(DPL)中的一部分地址进行比较。该方法由图 2 中所示的系统来实现。

图 2 的系统含有一订单管理系统 100, 比如 SMARTS 系统,它包括一用来存储个人和/或公司的运输和/或提货地址的数据库 110,个人和/或公司已经发出订单或者准备接受订单;和一个数据输入设备 120,它使用拼音字符以便将数据输入至数据库 110。仅仅显示了一个数据输入装置 120,但是实际应用中可能有多个这样的单元。

该系统还包括一第二数据库 130, 它用来存储英文语言的 DPL。

该系统另外包括一第一转换单元 140,它用来将该第一数据库 110 中的简体汉字数据项目转换成拼音数据项目,以便形成第一拼音数据库 150。该过程不删除数据库 120。

该系统仍包括一第二转换单元 160, 用来将该第二数据库 130 中的 英文语言数据转换成拼音数据项目, 拼音数据项目在第二拼音数据库 170 中。这个过程不删除该第二数据库 130。

最后,该系统包括一比较单元 180,它用来比较该第一和第二数据库 150,170 中的拼音项,以及一输出单元 190,用来将该第一和第二拼音数据库 150,170 中的各项目之间的任何相符合的情况通知给该系统的操作员,相符合的情况由比较单元 180 来完成发现。

图 1 的方法的头两个步骤(即,图 1 中短虚线以上的部分)是大家所熟知的将数据输入至该订单管理系统 100 的该第一数据库 110 中的步骤。特别地,在步骤 10 中,比如说销售代表中的用户使用数据输入装置 120 以便将诸如订货和运输地址这样的数据输入至订单管理系统 100 中。

13

图 3 中所示为由该订单管理系统 100 所产生的显示给用户的窗口。使用这个窗口,在步骤 20 中,通过用户的干预,该订单管理系统 100 将输入数据转换成简化的汉字双字节字符,以形成该第一数据库 110 中的项目。当从该第一数据库 110 中输出的项目被打印出来的时候它们是简体汉字的形式,通常被需要使用这些在运输和订货单的文档中。图 4 显示了从该第二数据库中输出的一个窗口,它具有用双字节简体汉字字符书写的全部的订货和邮寄地址。注意,数据库 110 有可能更进一步包含与中文无关的,并且甚至与本发明不相关的项目。这样的项目,如果它们已经是用英文语言的方式,有可能通过众所周知的方法直接地与数据库 130 中的项目(例如,非中文项目)进行比较。

在步骤 30 中,以简化的汉字双字节形式驻留在该第一数据库 110 中的订货和运输数据,由该第一转换单元 140 将它们转换成拼音字符,以便形成该第一拼音数据库 150 中的项目。如上所述,一单个的简体汉字字符可能对应多个拼音字符集。因此,对于每个该第一数据库 110 中的简体汉字项目,该第一转换单元 140 将生成所有可能的拼音字符集,它们能够由所述的项目派生出。而且这些拼音字符集中的每一个在数据库 150 中都形成一个项目。尽管如此,我们已经确认这种"简化的"过程不会损害鉴别过程的完整性。

特别地,在步骤 30 中由转换单元 140 完成的转换可以通过使用一个转换文件来实现,该转换文件比如像已经装载的微软视窗 98 的简体中文操作系统的缺省复制,对于每一个安装的该缺省文件系统的位置均能够在每一台 PC 机的 C:\Windows\system\winpy.com 处找到,其中在每个 PC 机里都安装了这种操作系统。

图 5 显示了步骤 30 的过程的一个例子。图 4 的窗口中所显示的地址是该第一数据库中的第 4602249011 号订单,如图 5(a)中所示。图 5(b) 所示为能够将每个简体汉字字符转换成拼音的各种方法。大多数字符仅仅只有一个拼音译文,可是他们之中的三个具有两个拼音拼写,这两个拼音拼写中的一个用阴影显示出。

使用图 5(b)的表格,简体汉字字符的字符串被转换成拼音字符的字符串。每个具有多个拼音表示符的简体汉字字符被转换成一个表示符后紧接着另外的表示符。通过指明对应于每个这样的汉字字符的其

后紧接着另外的带阴影的拼音表示符的第一拼音表示符,这些字符串被显示于图 5(c)中。

在步骤 40 中,通过该第二转换单元 160,该第二数据库 130 中的中文地址被转换成拼音以形成该第二拼音数据库 160 的项目。注意,该转换过程必须由一个能说中文的操作者手工地正常地去完成,尽管在原则上该过程也可以自动地或半自动地被实现。

表1显示了这种转换操作。表1显示了图1的方法中由DPL表所产生的拼音字符的数据库。每一行对应于DPL列表中的一个实体(标为PIN_YIN_1至PIN_YIN_9),例如,实体PIN_YIN_2是"Beijing Institute of Structure and Environmental Engineering(北京建筑和环境工程学院)"。第二方DPL表包括一对应于该实体的"No.36 Wanyuan Road Beijing China (PRC) (中国北京万园路 36号)"的地址(在表1中该地址被标为"BXA DPL地址")。注意,该地址是一个常规的英文文字(如,"路")和拼音(如,"万园")的混合体。在步骤40中,该BXA DPL地址被转换成(比如,由一个操作员)一个完整的拼音地址。作为参考,相应的简体中文地址显示于表1的右手栏中,尽管,本发明并不需要产生这一栏目。

尽管在原理上,将 DPL 列表中的所有项目都转换成拼音是有可能的,但本实施例仅仅将 DPL 列表中的中文项目的地址进行转换,例如,在本上下文中的"中文"可以被定义成这样的项目,它们是中华人民共和国境内以及其他可选领土之内的地址。通过采用这种"简化的"的手段,转换的数目(因而同样对于紧接下来的比较)被大大地减少了。一般地来讲,由于鉴别过程基于地址的基础之上,而且地址的属性并没有发生"转变",上述方法不会降低鉴别的完整性。

在步骤 50 中,对于该第一和第二拼音数据库 150,170 进行比较,以便确定它们是否相符合。这些是通过自动地抽取该第一数据库中的拼音字符串(例如,图 5(c)中所示的字符串)与该第二数据库中的拼音字符串(表 1 的"拼音地址"栏目)之间的相符合情况来完成的。

图 6 所示为一可以选择地显示给用户的窗口,通过比较单元 180 以便让用户确定这些匹配是怎样被处理的。如图所示,已经发现了一个在订单号第 402211081(如图 4 和 5,以及图下的上半部分所示)与

表 1 表中的实体 PIN_YIN_4 (如图 6 的下半部分所示)之后的可能相匹配的情况。注意,DPL 列表中的实体名称("BEIJING AEROSPACE AUTOMATIC CONTROL INSTITUTE (北京航空自动控制有限公司)")与名称("DALi Furniture (China) Ltd.)(大理家具(中国)有限公司)")是不相同的,其中应用名称(大理家具(中国)有限公司)来制作订单;本实施例已经发现这种相符合情况只是单单基于地址的基础之上。通过向图 6 窗口中的适当的可选对话框中输入点击,然后点击"OK",用户可以指示这种相符合情况是怎样被处理的。

, (

如果愿意的话,步骤 50 可以由运行该订单管理系统的机构的 DPL 遵循部门来完成,该相符合的情况可以与地方的 DPL 列表相合作,即,运行该订单管理系统的机构将发展同表中所列的各方(可不同于发展方列表)进行交易,至少是没有筛选鉴别操作。紧接着,地方 DPL 列表有可能会被使用以便附加到出口管理系统中,用于出口法规的遵循鉴别的目的,这正如同出口/运输文件的产生一样。

因此,步骤 30 和 40 已经形成了一公共平台(拼音),这样在步骤 50 中即能够实现中国订单的地址的遵从一致的鉴别。

该实施例可以采用在处理的方式进行操作,其中该第一数据库 110 中的多个项目(比如,该第一数据库 110 中的所有中文项目)一个接一个地(例如,像一连续不断的序列)被转换成拼音项目,以便形成该数据库存 50,然后数据库 150 中的每一个被转换项目均与该第二数据库 170 的被转换项目进行比较(比如说,一个接一个地)。

可以选择地,对于该第一数据库 110 的项目可以分别地进行步骤 30 (例如,每当有一新的项目被添加至该第一数据库 110 中),而通过将每个被转换的项目与该第二拼音数据库 170 的所有被转换项目进行比较,且为了数据库 150 中的所得项目可以进行步骤 50。如果没有发现相符合的情况,该数据库 150 的内容可以舍弃掉,换句话说,在本实施例的变化中,该第一拼音数据库 150 在任何时候不需要含有比由数据库 110 中的简体汉字项目中的某一单个项目而派生出来的拼音项目的数目更多的拼音项目数目。

步骤 50 中的比较可以用如上所述的方法来进行。如果有任何一个 匹配情况被发现,即要使用输出单元 190 以便通知该系统的操作员,

他(她)可以对相应订单进行特殊处理。

表 1

ID	名称	BXA DPL 地址	拼音地址	简体中文地址
PIN_YIN	ABC	Jianguomenwai Diplomatic Housing Compound, 8ldg 7-1 5th FL Apartment 4 BEIJIN , CHINA (PRC)	POU BU ZHU CHE ZHAJ GOUQU 7-1, 5 LOU 4 CHANDANYUAN	中国北京建国门外外 交部住宅区7-1, 5楼4单元
2 .	OF STRENGTH AND ENVIRONMENTAL ENGINEERING	No. 36 Wanyuan Road BEIJING , CHINA (PRC)	ZHONGGUO BEIJING WAN YUAN LUOLU 36 HAO	中国北京万园路36号
_3 _	OF STRUCTURE AND ENVIRONMENTAL ENGINEERING	No. 36 Wanyuan Road BEIJING , CHINA (PRC)	ZHONGGUO BEIJING WAN YUAN LUOLU 36 HAO	中国北京万园路36号
	AEROSPACE AUTOMATIC CONTROL NSTITUTE	BEIJING , CHINA (PRC)	ZHONGGUO BEIJING YONG DING LUOLU 51 HAO	
_5	ACADEMY OF SPACEFLIGHT	Road SHANGHAI , CHINA (PRC)	HANG TIAN DAIDA XIASHA	航天大 <u>度</u>
PIN_YIN 5 _6	SHANGHAI NSTITUTE OF	388 Cang Wu Rd.	CHUANG CANG WU LUOLU 388 XIAOHAO	中国上海仓武器388号
PIN_YIN_ _7	BCD	Housing Compound Bldg. 7-1 5th Floor Apartment 4 BELLING CHINA (PRC)	GUO MEN WAI WAI JIAO POU BU ZHU CHE ZHAI GOUQU 7-1, 5 LOU 4 CHANDANYUAN	中国北京赴国门外外 交部住宅区7-1。 5楼4单元
NIY_NIC 8	PEF	2-8-1 	BANGONG DAIDALOU 2-8-1	办公大模2-8-1专
9 PIN_YIN	FG H	Jnit 08-06/07, Liang Ma Fower North Dong San Huan Rd., Chao Yang District	ZHAOCHAOYANG GOUQU	中国北京朝阳区东三 环北路8号亮马大度08 -06/07

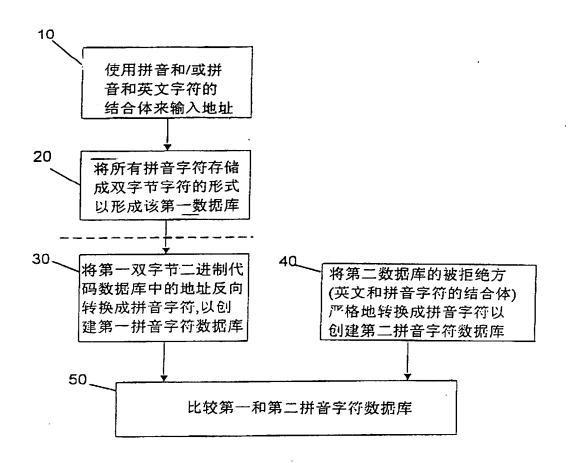


图 1

1

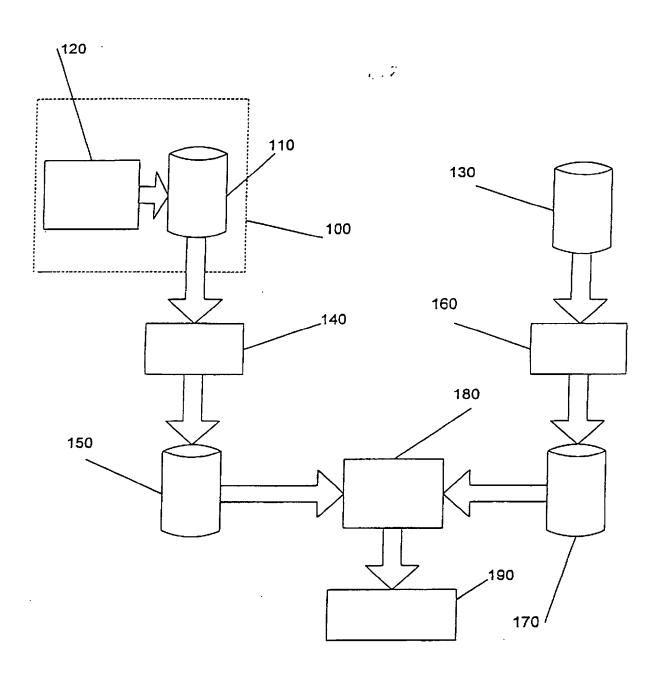


图 2

ſ	f
	1

Quotation: Cus		Add/Edis/Delate Questation of Standard Questation 紀次初記録	af dec	
Address Details Billing Address Address	Shippir	ng Address	Mailing Address	
	State : BJD go Code (RV DA : Zip Code : Country : CN	Merge Code : &/	State: OA: Zip Code: County: CN	
Billing Address In Nome Address beijing Address:		Mailing Addr Name: Address: Address: Tel:	ess in Double Byte	
Fin Code: Shipping Address Name: Address:	4:北京电视台diens in 5:北京地区dieu — 6:背井离乡lixia — 7:北京人ren	Fin Code : Tel Nos :	Fax Nos :	-
Address: Tel: Pin Code:	- 8:北京市shi - 9:北京时间shi ji 0:北京站zhan	Esil		

图 3

	w Quotatums/Orders) Jock Beld Bocard Quary Y 4 E		iter China Corp		scarp>oerr \$5 oct
Address	Details illing Address		Address	Mailing Address	
State: 8 Zip Code: Country: C	···	State: BJ000 Zp Code: CN	THE S		<u></u>
	illing Address In Double 比如传递这个写话51号	Byta	Mailing A	ddress in Double Byte 区外运路51号	
	sipping Address In Doub 比京河流及永宝路引导	ile Byte	Pin Code: Tel Nas:	Fox N	09:
Pin Code:	ICHNIES/KEH31*				
rn cam:				<u> </u>	

图 4

V

Order: 4602249011

			_						
16	京	海	淀	X.	旅	定	路	51	몽

(a)

双字节字符	拼音转换的所有可能性
北	bei
京	jing
海	hai
淀	dian
X	gou; qu
永	yong
定	ding
路	luo; lu
51	51
岁	xiao; hao

(b)

bei jing hai dian gou qu yong ding luo lu 51 xiao hao

(c)

图 5

ptnr_id: name_i: address_i:	es are pos : 46022110 : DaLi Fur : bei jing	81 niture (Ch hai dian	ina) Ltd. gou qu yo	
eviously un	nscreened.	Is there	a valid	match for this
			•	
				□ Update
	<u> </u>			
				_
			ſĔ	
			· · · · · · · · · · · · · · · · · · ·	•
-			OK	Help
			EROSPACE AUTOMATIC CONTROL INSTITUT BEIJING YONG DING LUCLU SI HAO	·

图 6